Technical Working Group, II

Teruko Manabe
Steven Worley
Miroslaw Mietus
Shawn Smith
Simon Tett
Volker Wagner
Scott Woodruff
David Berry
Liz Kent

NCAR
Scientific Computing Division

Introduction / Mission

The current digitization and archive rescue taking place around the world are making a significant contribution to the global marine surface data collection. These data are critical for a wide variety of climate studies, model initializations, and verification for remotely sensed measurements.

Observations held in collections from meteorological or oceanographic services, navies, and merchant marine sources are all valuable

NCAR Scientific Computing Division

Introduction / Mission

Every effort should be made to find more digital and non-digital data. All data not included in current archives are important. However, data prior to 1960, data from high southern latitudes, data from inland seas and lakes, and any information specifically from the two world war periods would be especially beneficial. These regions and periods are currently under represented in the archives.

Metadata coincident with the observations are very important. Independent metadata regarding instrument types and positions on ships and buoys are essential. Moreover, requirements included metadata describing shipping fleets, observational procedures and methods, and coding schemes.

NCAR
Scientific Computing Division

Recommendations

We recommend that data to be selected for digitization are those which add most information to the growing *IDATA* database. Rescue of data from all aging media is vital to prevent permanent loss of important archives. WMO committees should enhance the search effort by promoting world wide surveys

- Appendices attached show the current data sources in *IDATA,* data sources that are known to be available, and a time series of annual numbers of reports.
- Figure x shows the improved decadal station distribution resultant from data rescue and digitization currently in *IDATA*. Figure y is an example of data to be included and shows the station distribution of the Japanese Whaling Data.

NCAR
Scientific Computing Division

Recommendations

We recommend providing access to preliminary or interim data prior to inclusion in *IDATA.* This will better serve the research community with more current and all available data.

- Version and product identification for the following data categories should be added to all data records to clearly distinguish *IDATA* and its interim products. This information is vital to track which data have been used for research.
  - *IDATA*
  - Real-time ship and buoy data from GTS
  - Delayed mode data from the Global Collection Centers
  - Newly digitized or rescued digital data
- Clear documentation must be posted to advise user of potential duplication and lack of QC in the interim products.

NCAR
Scientific Computing Division

Recommendations

We recommend that a new format be used for the interim data and when possible for newly digitized or transcribed data.

- This will facilitate data exchanges and decrease the update cycle for *IDATA*.
- The IMMA format being develop under JCOMM is a candidate and should be completely and accurately documented.
- Software support in the form of example codes and Åfrequently asked questionsÅ should be made readily available.

NCAR
Scientific Computing Division

Recommendations

We recommend that:

the real-time data collection centers (e.g. MO, DWD, NCDC) keep original copies of the GTS data stream,

NCDC serve as the source for real-time interim data for *IDATA,*

and a comparison of GTS receipts at these collection centers be made to insure archive integrity.

- We recommend a pilot project to compare monthly files and expect one year of archive comparison to be sufficient to insure integrity of a single source for *IDATA*.
- A comparison format is to be determined by mutual agreement, IMMA is a candidate for consideration.

NCAR Scientific Computing Division

Recommendations

Concerning global blended land and ocean surface data sets:

We recommend this NOT be part of the technical tasks related to marine surface database development.

- Database development should focus on providing quality controlled uniformly formatted marine collections.
- Summaries of the observations should be provided as standard products.
- Analysis experts are required to create full global analyses

NCAR
Scientific Computing Division

Recommendations

We recommend including modern high quality data in *IDATA*.

- Include data at a higher observational frequency than standard synoptic periods, e.g. WOCE surface meteorological archive and data from other past and forthcoming national and international research programs.
- Near surface measurements from oceanographic collections received at NODC, e.g. ARGO
- Measurements from new automated VOS systems as they become available, e.g. VOSClim, VOS IMET

NCAR
Scientific Computing Division

Recommendations

We recommend at least one mirror data site.

- The mirror insures IDATA security and safe guards against accidental loss.
- The mirror should be located such to facilitate easy distribution worldwide.

NCAR Scientific Computing Division

Recommendations

We recommend continued development and application of new QC techniques and utilization of metadata.

- Systematic inclusion of platform metadata should be implemented as a part of *IDATA* observational records. The adaptive capability of the IMMA format may be suitable.
- New time varying statistical QC is required to improve the observational archive.
- Cross comparison of observations with advanced analyses could further enhance the detection of outlier data.

NCAR
Scientific Computing Division

Recommendations

We recommend continued wide distribution of all data in appropriate formats, and encourage sharing of software to access and analyze the data.

We recommend data should be freely available or available at a minimum cost for media.

NCAR
Scientific Computing Division